LA-UR-22-22884

Approved for public release; distribution is unlimited.

Title:	Ontologies: The Gateway to Knowledge-Enabled Information Services at Los Alamos National Laboratory
Author(s):	Sims, Benjamin Hayden; Blackhart, Craig; Pope, Paul Albert; Gordon, Andrew Lloyd; Turner, Andie Louise; Overmyer, Trinity Celeste; Gotches, Kimberly M.
Intended for:	Report
Issued:	2022-03-28

Los Alamos NATIONAL LABORATORY

Ontologies: The Gateway to Knowledge-Enabled Information Services at Los Alamos National Laboratory

March 23, 2022

Project team

Leadership
Andie Turner
Andrew Gordon

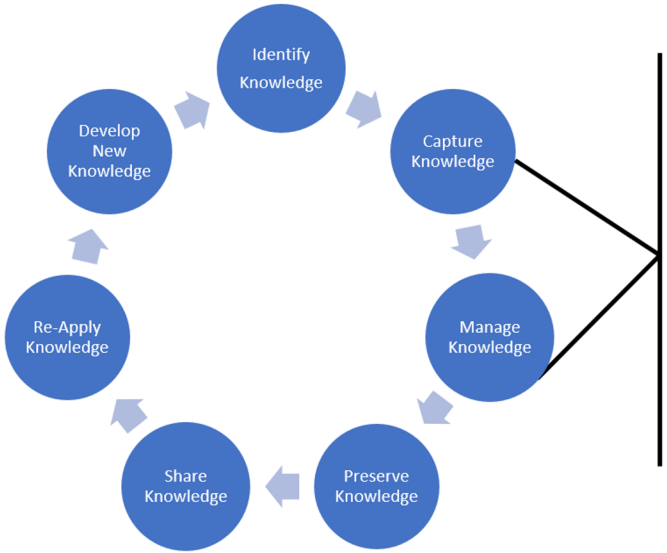
Ontology development
Benjamin Sims
Craig Blackhart
Paul Pope

KM support
Trinity Overmyer
Kimberly Gotches

Knowledge Management

Why Knowledge Management?
Nuclear weapons have and will continue to play a critical role in nuclear deterrence. Los Alamos National Laboratory has been preserving and transferring weapons knowledge since the Manhattan Project. Ongoing knowledge capture, preservation and transfer is necessary to maintain the capabilities needed for weapons innovation, responsiveness, and mission delivery – now and in the future.

What is Knowledge Management?
Knowledge management (KM) is a multidisciplinary field that reduces the risk of knowledge loss and promotes knowledge retention and sharing through systematic “knowledge cycle” activities.



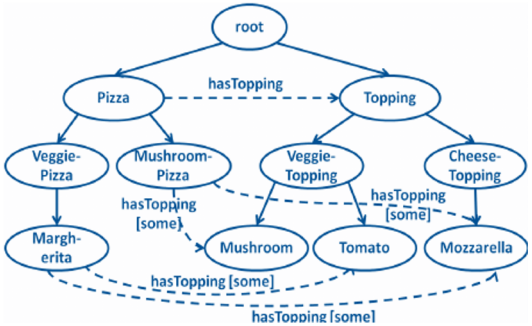
Ontology Captures Knowledge & Enables Knowledge Organization and Discoverability

- **Ontology** – technical knowledge domain mapping into a machine-readable format for enhancing search capabilities
- **Process** – standardize knowledge elicitation and documentation processes of ontology building and engineering
- **Integration** – share knowledge capture processes and tools within and across organizations as a part of business practices

What we will be covering today

1. Background
2. Introduction to ontologies
3. Motivating challenges
4. LANL ontology efforts
5. Ontology applications: Titan on the Red

At LANL, we use ontologies to capture and maintain essential organizational knowledge and support tools and frameworks for information discovery



A unique body of **knowledge** is at the core of everything we do.

Our libraries and archives house enormous amounts of **information** that must be not only **accessible** but **discoverable.**

We can't improve the way people get knowledge ***out*** of our libraries and archives without building more knowledge ***into*** our search tools.

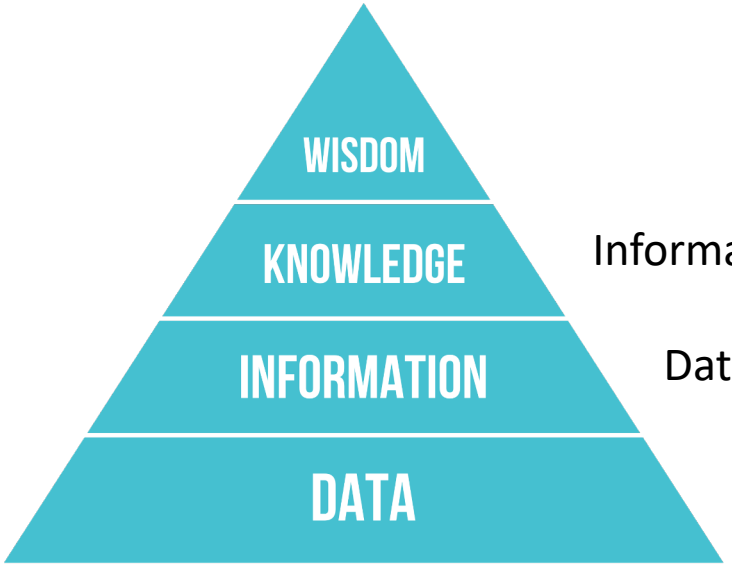
Ontologies capture **knowledge** as networks of **concepts** linked by **relationships** and **constraints**

Ontologies capture knowledge in **machine-readable formats** that provide **meaningful structures** for interpreting **information**.

Ontologies are a **gateway** to **knowledge management**, **enhanced search**, and **information discovery.**

Knowledge, information, and ontologies

Ontologies capture **knowledge,** creating **meaningful structures** for finding and interpreting **information**



Information in a human **context** that provides **meaning**

Data that is **organized** and **communicated**

Introduction to Ontologies

Example: some text …

Experiment number 25 used a 22 kg Type 202 vessel. The loading was conducted at 250C and 10 kPa, but was terminated due to anomalous readings from the Mech 15C. The X mode was set to 2A, so the source of the anomaly is unknown.

How a non-expert (or basic text analysis) might see it …

Context: Experiment?

Is this a type of vessel? Is it a shape? A function?

What is *kPa*?

Is this a temperature?

Experiment number 25 used a 22 kg Type 202 vessel. The loading was conducted at 250C and 10 kPa, but was terminated due to anomalous readings from the Mech 15C. The X mode was set to 2A, so the source of the anomaly is unknown.

Is *this* a temperature?

What does 2A mean?

What is a *Mech*?

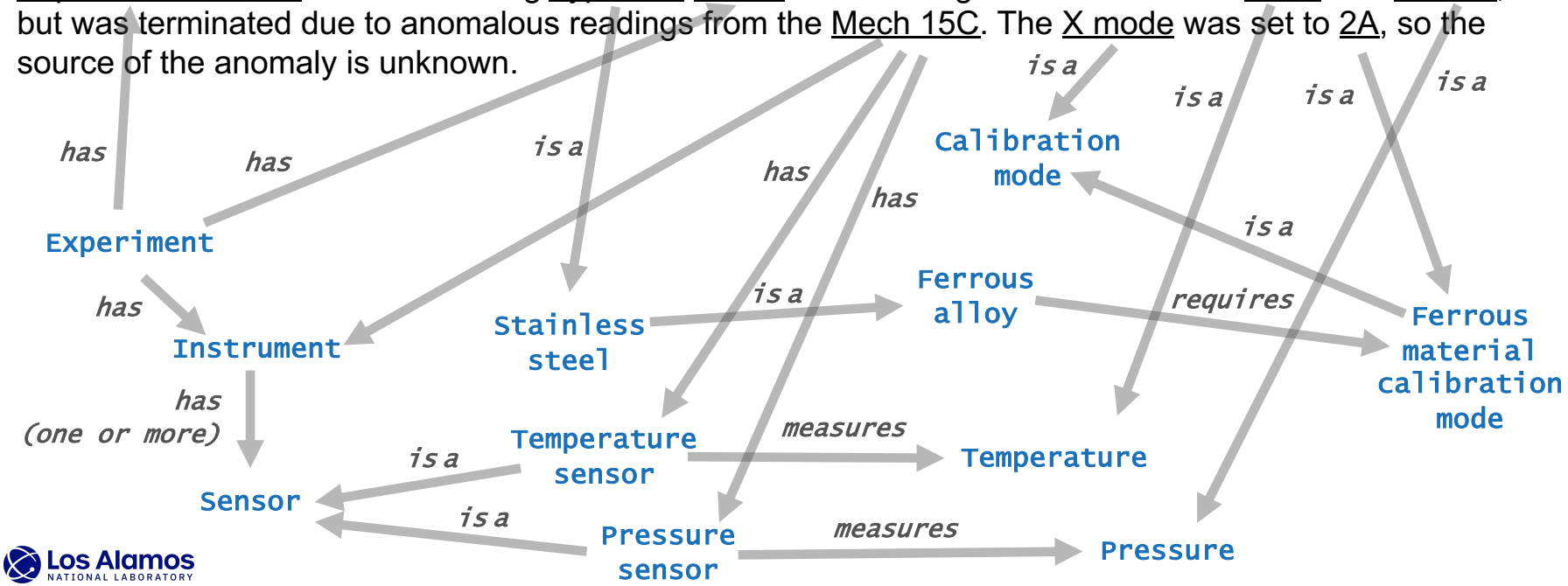
What is an X mode?

Why does X mode being set to 2A mean the source of the anomaly is unknown?

How an expert (or ontology-aided analysis) might see it …

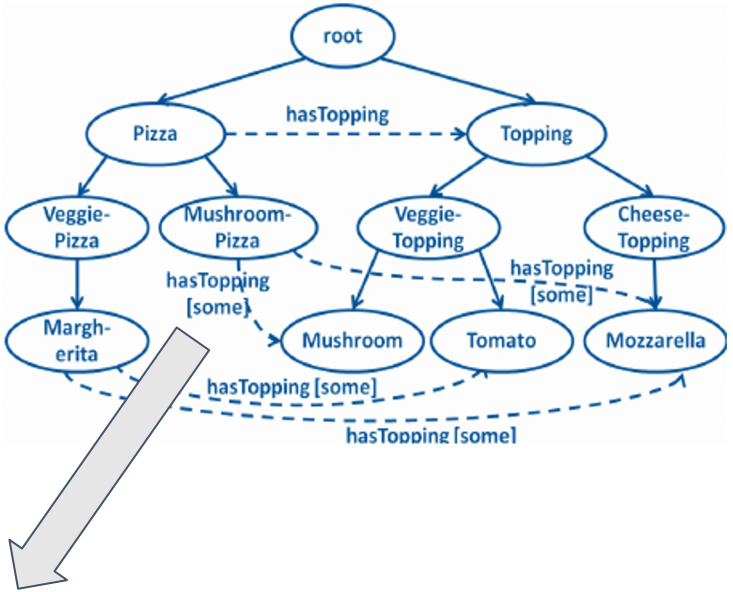
Context: Testing the suitability of different metal alloys for pressure vessels

Experiment number 25 used a 22 kg Type 202 vessel. The loading was conducted at 250C and 10 kPa, but was terminated due to anomalous readings from the Mech 15C. The X mode was set to 2A, so the source of the anomaly is unknown.



Example: Pizza Ontology

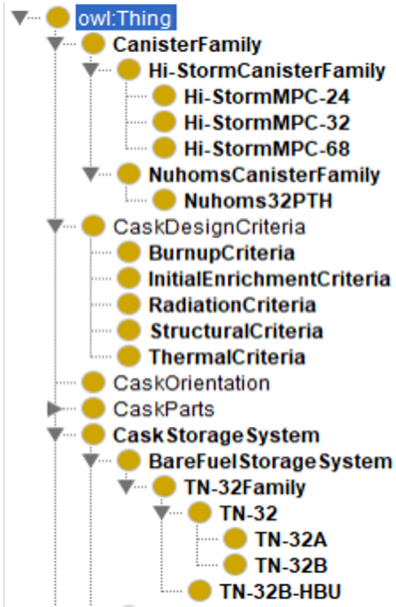
- Ovals contain **concepts** (classes)
- Solid arrows form a **taxonomy** (is-a/class-subclass relationships)
- Dotted lines represent additional **relationships** and **constraints**
- This creates a framework for describing real-world **individuals**



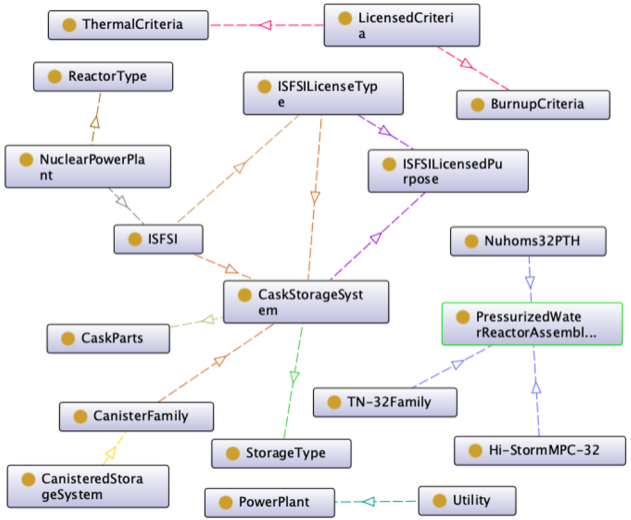

```
Ed's mushroom pizza  -  hasTopping  -   Smoked mozzarella
Ed's mushroom pizza  -  hasTopping  -  Portobello mushroom
Ed's mushroom pizza  -  hasTopping  -   Cremini mushroom
```

What a fully-developed ontology looks like (part of it)

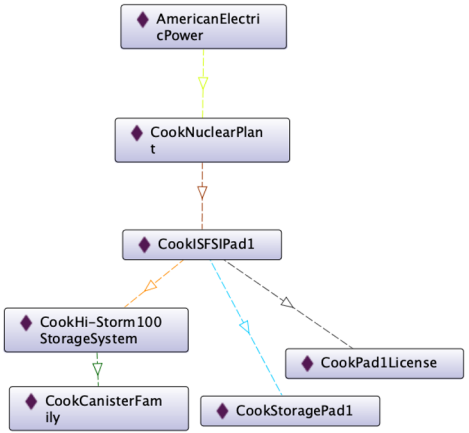
Taxonomy



Other relationships



Relationships applied to individuals



Why makes ontologies useful?

- **Condense information:** Efficiently distill essential concepts and relationships from years of SME experience and hundreds of documents
- **Reduce ambiguity:** Specify standardized terminology and equivalent terms
- **Enable knowledge management:** Explicitly capture knowledge and assumptions that might otherwise be lost or misinterpreted in the future
- **Enhance information resources:** Allow software to leverage human knowledge in context and make inferences, enhancing information search, retrieval, and analysis
- **Ready for integration:** Capture knowledge in standardized, non-proprietary, machine readable formats

Motivating Challenges

Why we need to leverage knowledge: Driving forces

- Archival challenges
  - Expanded use of multiple digital information repositories
  - Large-scale digitization of historical material
  - Need to make material discoverable, not just accessible
- Risk of knowledge loss
  - Retirement of the last generation of SMEs and weapons librarians with first-person knowledge of the nuclear testing era
- Rising user expectations for information retrieval
  - “Why can’t it be as easy as a Google search?”

Why we need to leverage knowledge: Unique challenges

- Technical community with unique, highly specialized knowledge base
- Need to leverage our entire organizational history to meet current challenges
- Changing mission (design/testing → science-based stockpile stewardship)
- Past: information discovery through personal knowledge of archives staff
- Present: growing archives team with increasing technological focus

1940s-1980s

1980s-2010s

2010s-

NATIONAL SECURITY RESEARCH CENTER

LANL NSRC Ontology Efforts

Two core ontologies under development

Weapons Design and Testing Ontology
- Nuclear tests
- Design features
- Connections between devices and tests

Started FY19, current version 1.0

Key connections

Components

Component properties

Institutions

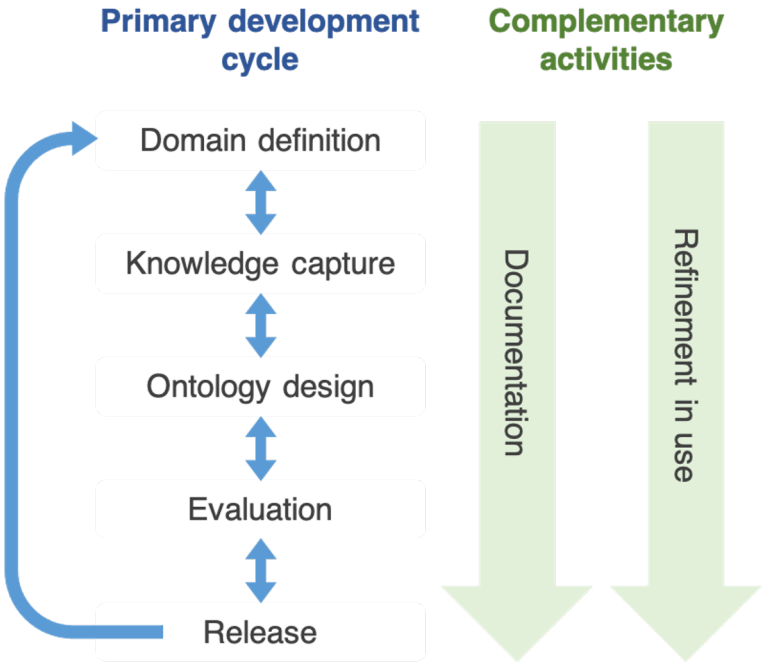
Etc.

Rocky Flats Collection Ontology
- Facilities
- Operations
- Manufacturing
- Processes

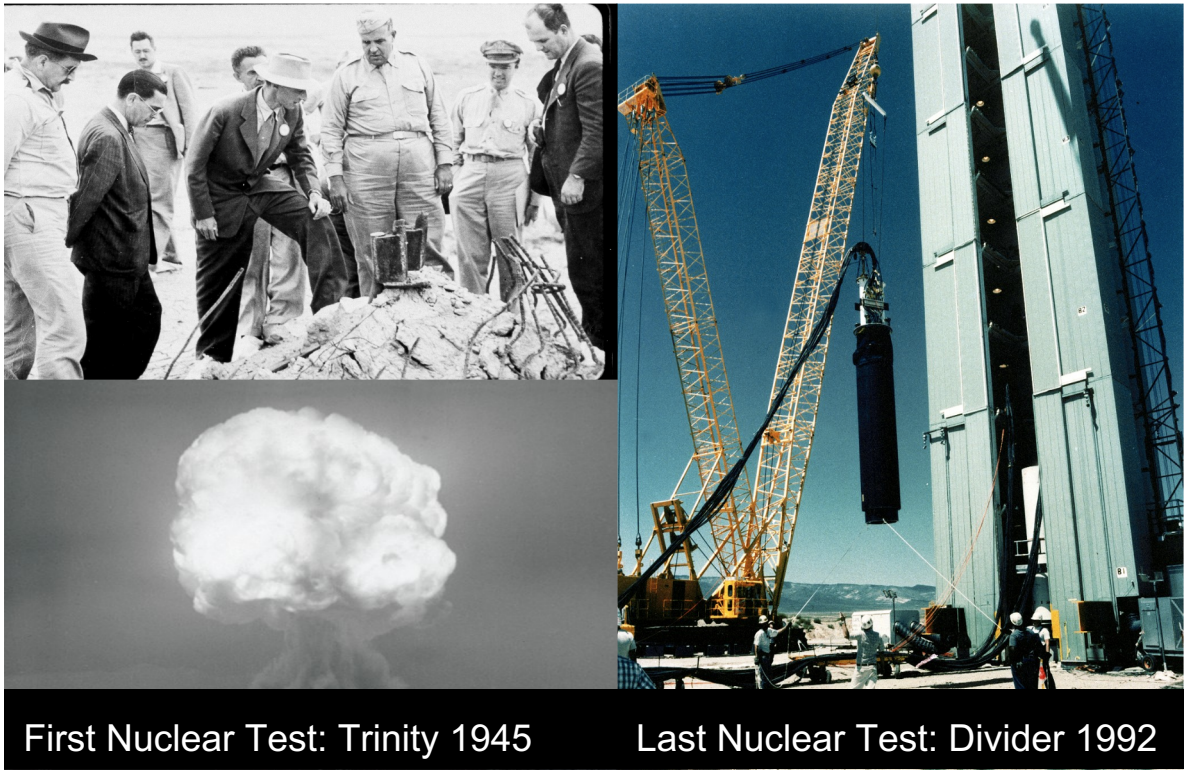
Started FY21, current version alpha

- Each covers several knowledge domains in the larger scope of weapons knowledge
- Designed to be interoperable with each other and future ontologies (modular development)

Creating and documenting a standardized ontology development process



Weapons Design and Testing Ontology



First Nuclear Test: Trinity 1945

Last Nuclear Test: Divider 1992

- **1054** tests performed over **42** years
- **30** years since last test performed
- **80** years of research
- **Goal:** Capture 80 years of knowledge of nuclear device designs and tests to enable intelligent search and other applications

Weapons Design and Testing Ontology

Scope (classified, version 1.0 completed FY21)

- **Nuclear weapons design** and **testing**, with a focus on the core **categories** and **relationships** weapons scientists most often use to navigate archival resources, including:
  - **design features** of **tested nuclear devices**,
  - **attributes** of historical **tests**, and
  - **associations** between **devices** and **tests.**
- Information sources:

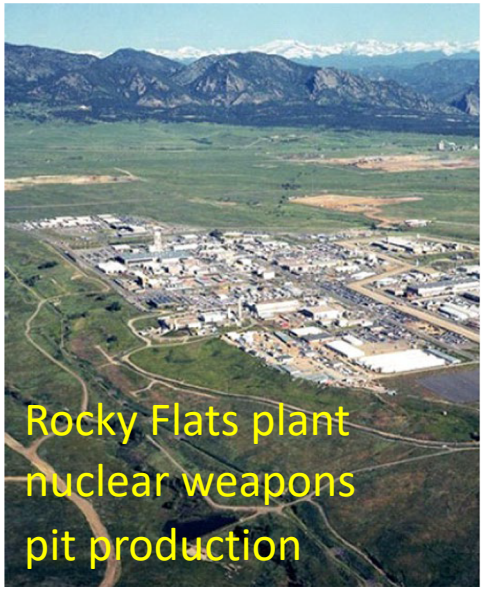
Community knowledge
- Design and testing SMEs
- Weapons librarians and archivists

Documents and data
- Paper and digital archives
- Legacy databases

External documentation
- OSTI references
- Unclassified documents
- Documents from other agencies

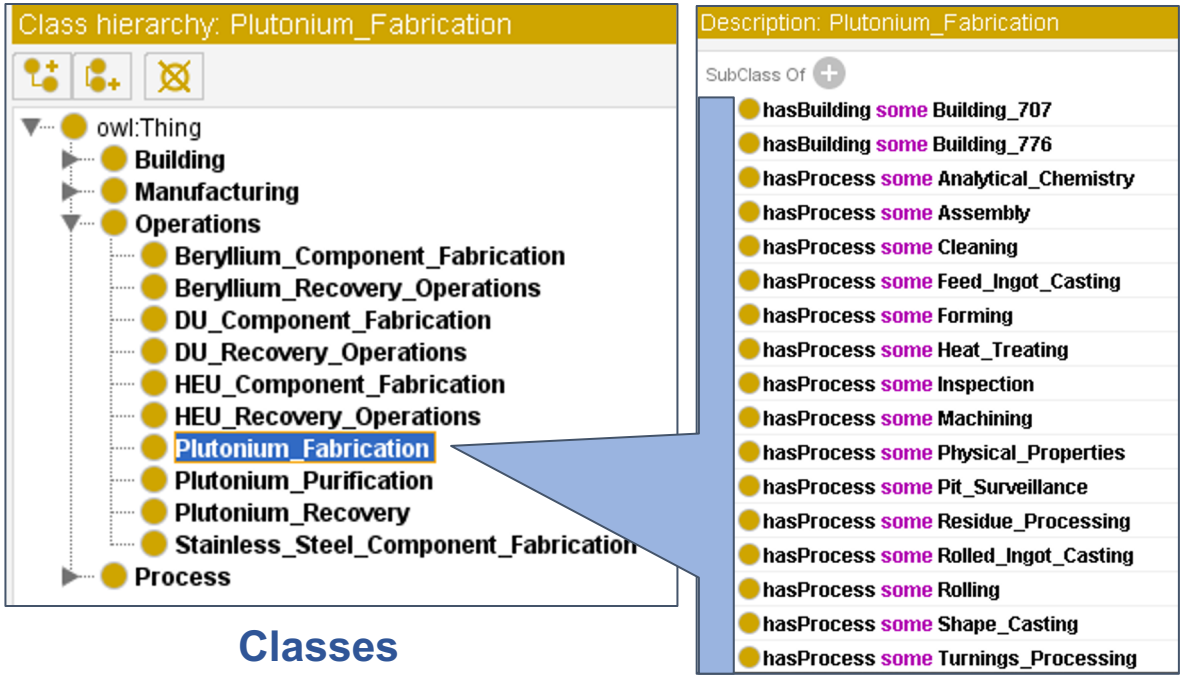
Archival structures
- Wellnitz collection structure
- Existing metadata categories
- Curated document folders
- Database schemas

Rocky Flats Collection Ontology



- **1989 – 2020:** thousands of documents stored at Denver Federal Center
- **August 2020:** transferred to LANL
- **Goal:** Enable a new generation of technicians, scientists, and engineers to better understand and carry out the process of pit production

Rocky Flats Collection Ontology



Classes

Class relationships

- Unclassified alpha version engineered and evaluated in FY21
- Unclassified high-level categories are the foundation of the ontology
- Classified version will graft more detailed information onto this foundation

Ontology applications: Titan on the Red

How ontologies support NSRC and Titan on the Red

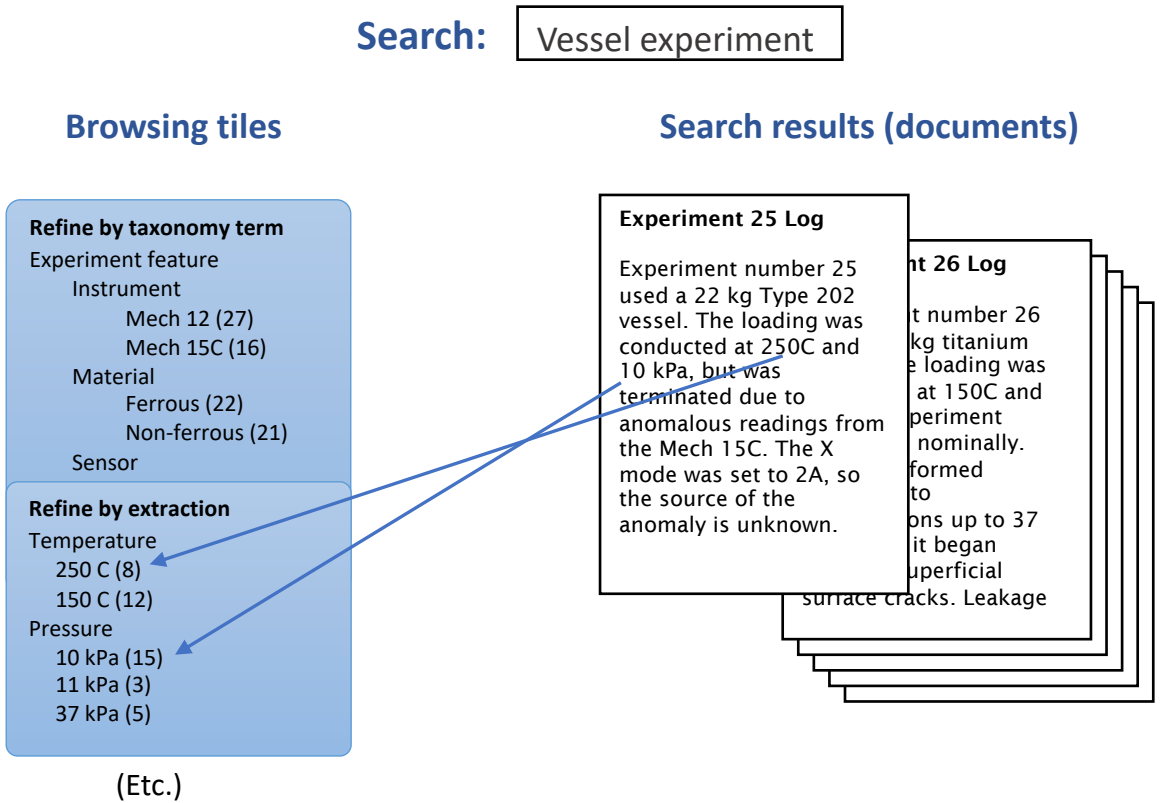
- Titan on the Red is a knowledge-enabled information discovery platform that will be the point of access for weapons information across multiple platforms
  - Incorporates linguistic technology to make sense of documents and terminology in context
  - Linguistic technology is driven by a general-purpose *knowledge graph* as well as domain-specific *taxonomies*
- Ontologies will be used to:
  - Build taxonomies and/or refine the knowledge graph to include LANL-specific topics
  - Generate linguistic rules for identifying LANL-specific concepts in text

How Titan on the Red leverages ontologies for information discovery

User interface
Search results (relevant documents)
Related taxonomy terms and associated documents
Extracted information (names, dates, measurements, etc.)

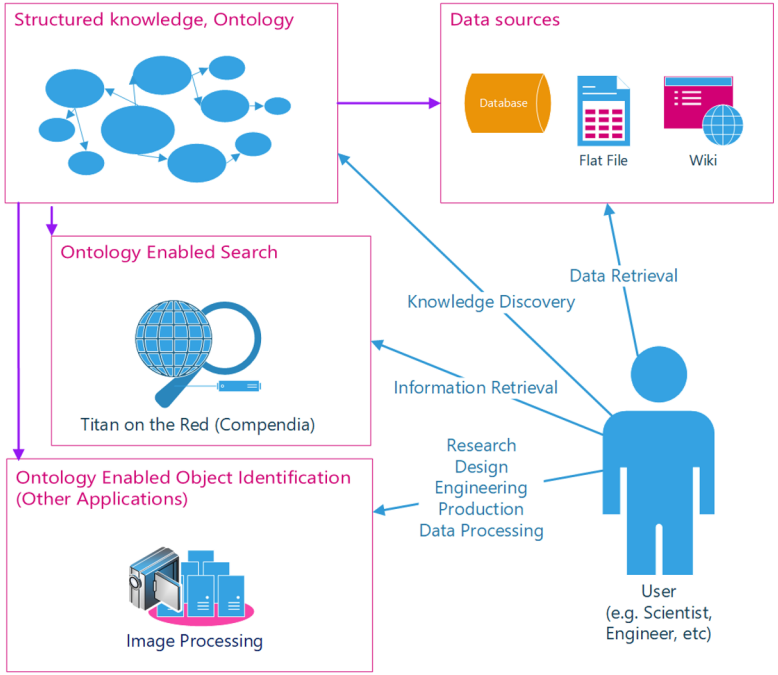
Ontology concepts and structures → Ontologists work with librarians and archivists to create application-specific taxonomies → Linguistic package developers create rules for identifying taxonomy topics in text → System uses rules along with built in knowledge graph to tag documents with relevant topics

How users will interact with Titan on the Red



Conclusion

- Ontologies capture **knowledge** in **machine-readable** formats, providing **meaningful structures** for interpreting **information.**
- At LANL, we are developing ontologies to support **knowledge management**, **information retrieval**, and **enhanced search tools** for **knowledge discovery**
- In the future, we plan to:
  - Continuously refine ontologies in use
  - Develop and integrate new ontology modules to expand scope
  - Explore additional ontology applications



We would like to hear from you!

Questions? Want to learn more?

Email us at wrs-ontologies@lanl.gov

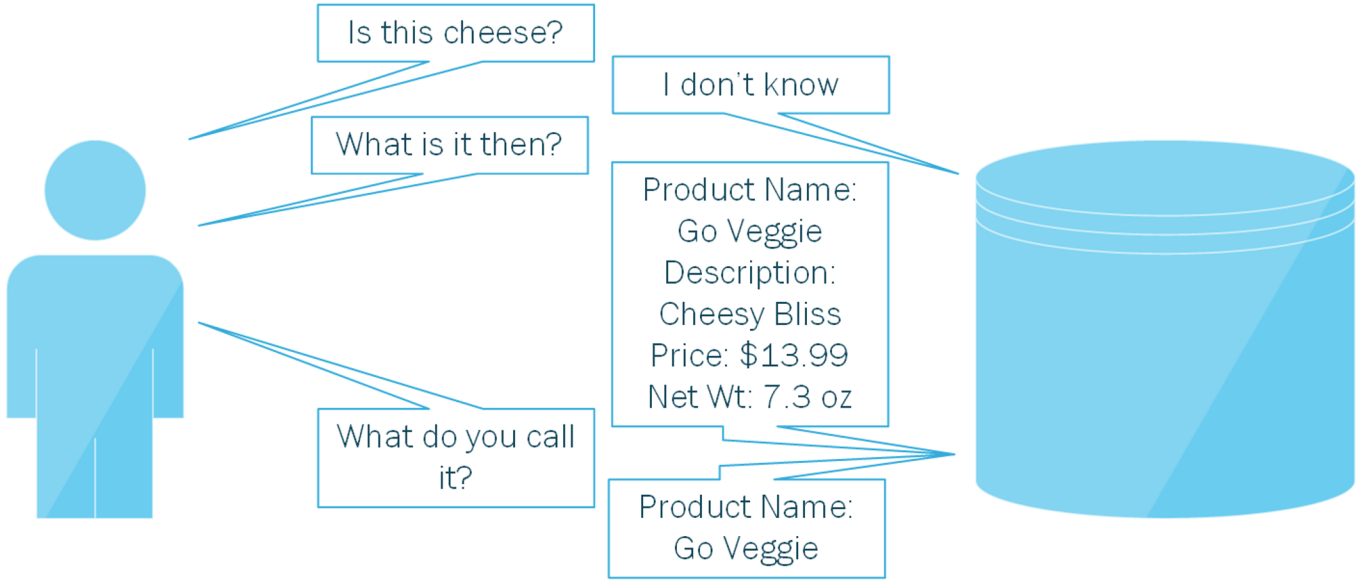
Backup Slides

What is an ontology?

An ontology is a **formal**, **explicit** specification of a **shared conceptualization**

- **Formal:** machine readable and based on formal logic
- **Explicit:** concepts, relationships and constraints are clearly defined
- **Shared:** captures knowledge of a group/domain
- **Conceptualization:** captures domain knowledge as a network of concepts

Questions a database can answer



Questions an ontology can answer

Is this cheese?

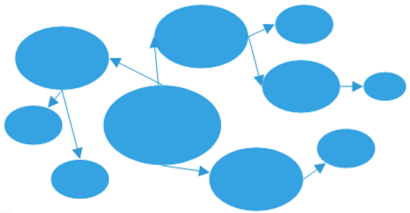
No

What is it then?

An edible, plant-based product marketed as a cheese substitute

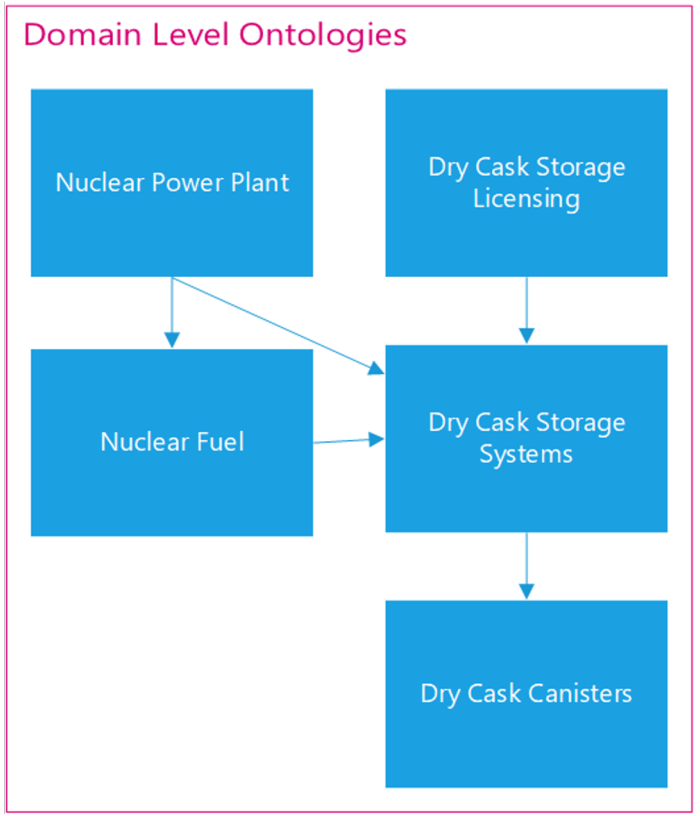
What do you call it?

Cheese alternative, vegan cheese, cheddar flavor block



Ontology engineering goals

- Meets requirements as laid out in competency questions
- Software enabling but not application-specific
- Modular
  - Expandable
  - Interoperable
  - Reusable
- Meets evaluation criteria



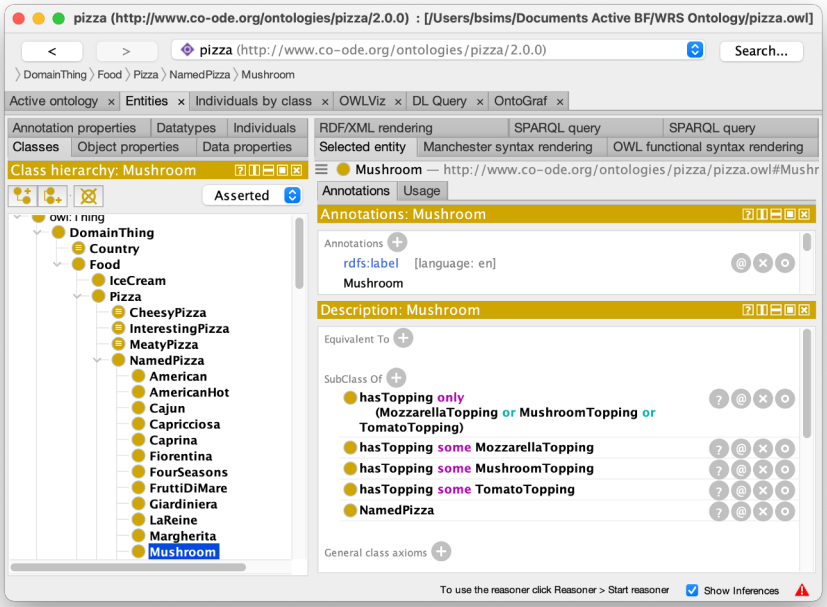
Metrics for evaluating ontologies

- **Completeness:** Does the ontology cover all relevant topics within scope?
- **Correctness:** Do terms and relationships accurately represent knowledge?
- **Logical consistency:** Are there any internal flaws in design or implementation?
- **Usability:** Is the ontology designed to support its intended uses/integrations?
- Challenges for evaluation
  - Underdetermination: No single correct representation (but many wrong ones)
  - Different perspectives: SMEs may disagree on how to categorize knowledge
  - Lack of ground truth: No definitive test or standard for true validation

Ontology development tools

- We develop ontologies using the Protégé ontology development environment
- Protégé saves ontologies in the W3C standard, machine-readable OWL description language

```
<owl:Class rdf:about="http://www.co-
ode.org/ontologies/pizza/pizza.owl#Mushroom">
    <rdfs:subClassOf rdf:resource="http://www.co-ode.org/ontologies/
    pizza/pizza.owl#NamedPizza"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="http://www.co-ode.org/ontologies/
        pizza/pizza.owl#hasTopping"/>
        <owl:someValuesFrom rdf:resource="http://www.co-
ode.org/ontologies/
        pizza/pizza.owl#MozzarellaTopping"/>
                                    . . .
```



Primary ontology development cycle

- **Domain definition**
  - Identify key documents and SMEs
  - Develop initial outline of topics and scope
  - Competency questions
- **Knowledge capture**
  - Read documents and work with SMEs to develop detailed list of concepts and entities to be included in the ontology
- **Ontology design**
  - Build structured outlines of ontology entities
  - Implement these outlines in Protégé
  - Develop further relationships and constraints among entities
- **Evaluation**
  - Assess ontology completeness, correctness and consistency with help from SMEs
  - Make final changes/refinements
- **Release**
  - Release ontology version; begin new cycle

Ontology project team

- Project leadership
  - **Andie Turner:** KM team lead
  - **Andrew Gordon:** NSRC librarian/archivist (project leader)
- Ontology development
  - **Benjamin Sims:** Sociology, knowledge capture
  - **Craig Blackhart:** Computer science, ontology engineering
  - **Paul Pope:** Geospatial sciences, ontology engineering
- KM support
  - **Trinity Overmyer:** Rhetoric and technical communication, knowledge capture
  - **Kimberly Gotches:** Knowledge systems and integration